

Hartmut Schmidt

Plädoyer für eine moderne korpusbasierte deutsche Wortschatzforschung¹

1. Ausgangslage

Politische Gemeinschaften bedürfen der verantwortlichen Prüfung ihrer Kommunikationssysteme, ihrer Kommunikationsmedien und Kommunikationsmittel. Ein erstrangiger Teilbereich dieser Prüfung sollte das Nachdenken über die Funktion der eigenen Sprache, über deren Entwicklung und über ihr Verhältnis zu anderen Sprachen sein. In der föderal konstituierten Bundesrepublik Deutschland ist Kultur Ländersache. Sprache wird in traditioneller Weise als Kulturgut begriffen, das – bis auf amtliche Eingriffe in die Orthographie – ganz überwiegend der Selbstregulierung unterliegt. Eine Instanz für die notwendige Reflexion über die Rolle des Deutschen, die sich auf einen staatlichen Auftrag, auf eine natürlich gewachsene Autorität oder auf den Konsens der wichtigsten institutionellen Interessenten stützen könnte, existiert nicht. Auch ein geordnetes Zusammenwirken mit anderen Ländern, in denen Deutsch Landessprache (oder eine der Landessprachen) ist, gelingt bisher nur in Teilbereichen (Orthographiereform, Ansätze zur naturwissenschaftlich-technischen Terminologiebildung).

Problemzonen, in denen die Mechanismen der Selbstregulierung derzeit nicht in der wünschenswerten Qualität funktionieren und für deren ungelöste Probleme zunächst wenigstens ein öffentliches Bewußtsein hergestellt werden sollte, sind z.B.:

(1) *Bürger und Staat*: Gesetzes- und Verwaltungssprache, Parlaments- und Politikerdeutsch; Verhältnis von politischer und sprachlicher Emanzipation; Hebung der aktiven und passiven Sprachkompetenz in der Schule und in allen höheren Ausbildungsgängen.

(2) *Alltagssprache und Fachsprache*: Kommunikation innerhalb wissenschaftlicher und technischer Diskursgemeinschaften; Vermittlungsinstanzen zur Allgemeinsprache (Werbung, Wissenschaftsjournalismus, Gebrauchsanweisungen, Terminologienormung und Terminologieberatung).

¹ Der Artikel beruht auf einer Ausarbeitung, die der Verfasser in seiner Funktion als Projektleiter der Berliner Arbeitsstelle des Deutschen Wörterbuchs am 15.2.1996 an den Präsidenten der Berlin-Brandenburgischen Akademie der Wissenschaften gerichtet hat.

(3) *Inländer und Ausländer*: Sprachliche Anforderungen an gelungene Integrationsprozesse. Spracherwerb und Einbürgerung. Sprachliche Folgen nicht gelingender Integration. Sprachlernkonzepte deutscher Schulen. Problemvergleich (Deutsch im Verhältnis zu anderen Sprachen) in Deutschland/Österreich/Schweiz/Italien/Belgien/Luxemburg/Frankreich usw.

(4) *Deutsch in der internationalen Sprachenbalance*: Sollte es eine wirksamere deutsche ›Sprach-Außenpolitik‹ geben (Deutsch in der EU, Deutsch in der UN, Deutsch im Internet)? Sollte Englisch (Französisch, Spanisch, Russisch usw.) ohne Vorbehalte zur Außenkommunikation der Deutschen akzeptiert werden? Soll und kann durch verbesserten Sprachunterricht wenigstens in Europa ein neuer Typus intersprachlicher Kommunikation gefördert werden (jeder spricht die Sprache, die er am besten kann; möglichst viele *verstehen* die akzeptierten Hauptsprachen)?

(5) *Sprachausgleich zwischen Staaten mit (auch) deutscher Landessprache* (Deutschland, Österreich, Schweiz, Italien, Belgien usw.) *und Sprachausgleich innerhalb Deutschlands*: Gibt es vernünftige Gründe für die Lockerung eines im Ganzen norddeutsch geprägten Normengefüges? Wird diese ›deutsche Standardsprache‹ heute zu Recht durch Vertreter des deutschen Südens, aber auch durch Österreicher, Schweizer (usw.) gelegentlich in Frage gestellt?

Das ist ein weites Feld, das hier nicht als Ganzes abgeschritten werden soll. Aber an vernünftigen Versuchen, die anstehenden Fragen genauer zu behandeln, sollten auch Sprachwissenschaftler interessiert sein (vgl. zuletzt Ulrich Ammon, *Die deutsche Sprache in Deutschland, Österreich und der Schweiz*, Berlin/New York 1995). Die meisten dieser Probleme können nur dann gründlich behandelt und verlässlich beurteilt werden, wenn verschiedene linguistische Disziplinen ihre Aufgaben erledigt haben: Dazu gehört, daß wissenschaftlich unanfechtbare Untersuchungen der Wortschatzentwicklung und lexikographisch einwandfreie Auskunftsmittel zur Verfügung stehen. Das gilt historisch und gegenwartssprachlich, also für die diachrone Erschließung dieses Wortschatzes, seine Funktion als Schlüssel zu deutschsprachigen Texten der Vergangenheit, als Instrument und Kode unseres historischen Gedächtnisses im Guten wie im Bösen, und ebenso für seine Funktion in den deutschsprachigen staatlichen oder regionalen Kommunikationsgemeinschaften der Gegenwart, für die Verständigung zwischen diesen Kommunikationsgemeinschaften und für alle Ebenen des intersprachlichen Austauschs. Nehmen wir diesen Fragenkomplex ernst, so müssen wir darüber reden, wie es derzeit um die wissenschaftliche historische Lexikographie des Deutschen steht (wobei »historisch« hier nur meint, daß der deutsche Wortgebrauch in seiner heutigen und seiner geschichtlichen Funktion gesehen werden sollte).

2. Defizite der deutschen historischen Lexikographie

Die wissenschaftliche historische deutsche Lexikographie befindet sich seit Jahrzehnten in einer zwiespältigen Lage. Zum einen genießt sie nationale und internationale Anerkennung für den hohen Standard wichtiger einschlägiger Leistungen wie des Deutschen Wörterbuchs von Jacob Grimm und Wilhelm Grimm (DWB), des Deutschen Rechtswörterbuchs, des Goethe-Wörterbuchs, des Althochdeutschen Wörterbuchs oder des Deutschen Fremdwörterbuchs. Zum anderen ist nicht zu übersehen, daß gerade ein Werk wie das Deutsche Wörterbuch in der Öffentlichkeit – und auch im Selbstverständnis der Disziplin – den Platz nicht mehr einnimmt, den es während seiner Erstbearbeitung bis in die Nachkriegszeit hinein wie selbstverständlich beanspruchen durfte. Das zeigt sich z.B. beim Vergleich der Mittel, die bis zur Beendigung der Erstausgabe im Jahr 1960 (Quellenverzeichnis 1971) eingesetzt werden konnten, mit den heutigen Möglichkeiten. Der Unterschied wird noch deutlicher, wenn wir auf die Potenzen und die Ergebnisse vergleichbarer europäischer Wörterbuchstellen schauen. Das französische Standardwerk für die Epoche 1789–1960 (*Trésor de la Langue Française*) erschien vollständig in 24 Jahren 1971–1994 in 16 Bänden, die meisten mit weit über 1000 Seiten; das englische Standardwerk (*Oxford English Dictionary*), das den gesamten Bezeugungszeitraum des Englischen erfaßt, erschien mit zwanzig fast ebenso starken Bänden in zweiter überarbeiteter Auflage in einem einzigen Jahr (1989), eine Leistung, die in Deutschland vorläufig undenkbar wäre.

Die Ursachen für das Zurückbleiben eines klassischen Leistungsangebots der deutschen Lexikographie hinter den Erwartungen liegen nicht in ausbleibendem Interesse an einschlägigen Auskünften (also bei der Öffentlichkeit) oder in mangelnder Leistungsbereitschaft der Mitarbeiter (also bei den Wörterbuchautoren). Problematisch sind natürlich die – bezogen auf die Erfordernisse – immer geringeren Haushaltsansätze. Als die eigentliche Schwachstelle erscheint gegenwärtig aber die trotz aller Fortschritte im Einzelnen immer noch wachsende Distanz zwischen der heutigen deutschen Wörterbuchpraxis und den technischen Möglichkeiten der internationalen Lexikographie.

Die Ursachen hierfür liegen in der seit Jahrzehnten fast schon als Tugend geltenden Genügsamkeit der Lexikographen, die sich daran gewöhnt haben, daß ihre Wünsche auf taube Ohren stoßen, und sie liegen in der gegenwärtigen Gleichgültigkeit der politischen und kulturellen Entscheidungsträger gegenüber sprachlichen Fragen. Weil Angelegenheiten der Sprache in der deutschen Vergangenheit immer wieder einmal zu nationalen Existenzfragen hochstilisiert wurden, fürchten wir heute, Geister der Vergangenheit wachzurufen, wenn wir öffentlich auf die Tatsache hinweisen, daß in den heutigen Staaten mit deutschsprachiger oder teilweise deutschsprachiger Bevölkerung Sprachfragen öffentliches Interesse und öffentliches Engagement verdienen.

Daß dieses Interesse beim großen Publikum latent vorhanden ist, zeigt z.B. die gegenwärtige Überreaktion auf einige Änderungen des gewohnten

Schriftbildes. Der Wunsch, genauere Auskünfte über Sprachgebrauch und Gebrauchsgeschichte zu erhalten, ist aber tiefer begründet als die Irritationen über eine geänderte Orthographie. Das öffentliche Nachschlage- und Auskunftsbefürfnis nach historischen Sprachdaten des Deutschen nimmt gegenwärtig trotz der sehr unterschiedlichen Leistungsfähigkeit der zur Verfügung stehenden Auskunftsmittel deutlich zu. Dafür sprechen die sich häufenden Neuauflagen und Neubearbeitungen wichtiger historischer etymologischer Wörterbücher. Dafür spricht auch die Tatsache, daß wohl alle Institutionen, die historische Wort- und Formulierungsforschung betreiben und z.B. über Internet und World Wide Web über ihre Möglichkeiten informieren, eine Vielzahl von Anfragen aus Fernseh-, Radio- und Zeitungsredaktionen, von Buchautoren, Werbetextern, Lehrern, Sekretariaten oder auch streitenden Parteien (z.B. im Zusammenhang von Patentkonflikten oder Beleidigungsklagen) nach dem Gebrauch und der Geschichte von Wörtern und Formulierungen erhalten. Aber keine der befragten Stellen verfügt – anders als die lexikographischen Zentralen in England oder Frankreich – über ausreichende Grundlagen für zuverlässige Antworten über den Gesamtbereich des die Fragesteller interessierenden bedeutsamen Wortschatzes deutscher historischer und gegenwartssprachlicher Texte. Ein Teil der Öffentlichkeit, der Medien und der Universitätskollegen hat sich leider daran gewöhnt, daß entsprechende Auskünfte in Deutschland nicht zu haben sind, und beschränkt sein Interesse auf die Einheiten des engeren Grundwortschatzes, die – einschließlich eines Kernbestands wichtiger Komposita und Ableitungen – in knappen und z.T. vorzüglichen historischen oder etymologischen Werken aufbereitet sind. Der trotz dieser Leistungen im ganzen mangelhafte Stand der lexikographischen Erfassung der historischen Materialien ist im übrigen eine wichtige Ursache auch für das unbefriedigende Niveau der linguistischen Annäherung an sprachgeschichtliche Problemstellungen.

Das große deutsche Standardwerk für die Behandlung wortgeschichtlicher Fragen ist das DWB. Zwar hat dieses Werk ein jahrzehntelanges Aschenputteldasein geführt, als der Verlag viele Bände über unzumutbare Zeiträume hinweg nur unvollendet in Lieferungen anbieten konnte, aber seit dem Abschluß in den Jahren 1960/1971 existiert der »Grimm« in 33 Lexikonbänden und seit seiner Auferstehung als dtv-Neudruck zielt er die Bücherschränke in überraschend hohen Auflagenzahlen. Sein Problem bleibt seine Entstehungsgeschichte. Zweimal wurde das Wörterbuch unter problematischen Rahmenbedingungen als nationales Unternehmen gefeiert: 1852 nach Erscheinen der ersten Lieferung und 1960 nach Fertigstellung der letzten. Aber nicht einmal die feiernde Nation hätte sich nach dieser Frist wiedererkannt. In seinen neueren und neuesten Abschnitten ist das Werk ganz gewiß eine vorzügliche lexikographische Leistung. Trotzdem dürfen wir nicht die Augen davor verschließen, daß das DWB in seinen 32 Bänden (A–Z) so inhomogen ist, daß es über weite Strecken versagt. Der größte Teil des Stichwortalphabets (A, B, C, D, E, F, H, I, J, K, L, M, N, O, P, Q, R) wurde ausschließlich im 19.

Jahrhundert bearbeitet. Diese Bände sind heute so vollständig und nützlich wie eine Bibliographie, ein Physikbuch oder eine Parteiengeschichte aus dem 19. Jahrhundert, vorzüglich in bestimmten Einzelheiten, unbrauchbar im Ganzen. Mit vertretbaren neueren Methoden wurden seit 1930 nur die *Abschluß*-teile der Buchstaben G, S, T, U, V, W, und Z sowie der kurze Abschnitt X und Y erarbeitet. Aber auch für alle neueren Teile gilt, daß die Materialgrundlage im wesentlichen in den Jahren um 1908 geschaffen wurde. Das bedeutet: Die Texte des 20. Jahrhunderts sind kaum oder doch nur in ungenügender Auswahl vertreten. In allen älteren Abschnitten bleibt das 20. Jahrhundert, inzwischen eine wichtige Epoche der deutschen Sprachgeschichte, für das Deutsche Wörterbuch ohnehin eine Blindstelle. Manche Ausnahmen, z.B. die Anführung ausgesuchter Texte der Nazizeit, sind oft eher peinlich als sachdienlich, weil ihnen die Einbettung in das Gesamtspektrum deutschsprachiger Texte ihrer Zeit fehlt. Welcher deutschsprachige Linguist oder Literaturwissenschaftler kann sich mit einem solchen Forschungs- und Interpretationsmittel zum eigenen Wortschatz zufrieden geben? Welcher deutsche Historiker findet sich heute noch im wichtigsten historischen Wörterbuch des Deutschen zurecht? Welcher französische oder englische Wissenschaftler, seit Jahren an die präzisen lexikographischen Instrumente des eigenen Landes gewöhnt, ist heute noch bereit, in der ersten (und einzigen abgeschlossenen) Auflage des DWB auf die Suche nach einer verlorenen Bedeutung zu gehen? Und welchem deutschen Publikum kann man es zumuten, die in einer Unzahl älterer DWB-Artikel dargebotenen wirr geordneten Ausschüttungen gelehrten Sammelfleißes nachzuvollziehen, in der vagen Hoffnung, vielleicht am Ende doch das Gesuchte zu finden?

Ein gutes Wörterbuch – gleich welcher Größenordnung – ermöglicht seinen regelmäßigen Benutzern den Aufbau einer festen Erwartungshaltung hinsichtlich der Auskünfte, die ihnen geboten oder nicht geboten werden. Wichtige Voraussetzungen hierfür sind die Einheitlichkeit der Anlage und Ausführung und die möglichste Knappheit der Bearbeitungszeit. Das DWB kann diese Erwartungen nicht erfüllen. Seine Benutzer müssen nicht nur die Abschlußdaten der Bände, sondern auch die Erscheinungsjahre der Lieferungen beachten, wenn sie der Materialgrundlage und dem Zeithorizont der Wörterbuchautoren gerecht werden wollen. Ein paar Beispiele aus einem einzigen Wörterbuch-Buchstaben mögen das belegen:

Wer etwas über das Stichwort *Garten* wissen will, wird mit dem Stand des Wortgebrauchs bis 1876, dem Erscheinungsjahr der betreffenden Lieferung, bedient. Der Artikel über das Wort *Greuelthaten* von 1919 verzeichnet als modernste Autoren Gutzkow und Schopenhauer. Der Artikel *Glaube* bietet 1939 als moderne Belege solche des Reichskanzlers Adolf Hitler (z.B.: *Der Glaube an den Sieg wich [bei den Alliierten] der Furcht vor der kommenden Niederlage*). Der Artikel *Götzendienst* bedient uns 1947 schon aus Thomas Manns *Dr. Faustus* und *grauenvoll* 1958 immerhin mit Feuchtwanger und Victor Klemperers LTI. Und jede dieser Lieferungen leistet das ihr Mögliche

mit den lexikographischen Methoden, den semantischen Konzepten und den gesellschaftlichen Grundüberzeugungen ihrer Zeit. Jede dieser Lieferungen ist deshalb nicht nur neutrales lexikographisches Auskunftsmittel, sondern Denkmal einer Epoche. In die Buchdeckel des DWB gebracht, entsteht so ein ziemlich widerspruchsvoller Mix.

Um das DWB den Ansprüchen einer neueren Zeit anzupassen, wurde in den fünfziger Jahren durch die Berliner und die Göttinger Akademie seine Teilneubearbeitung für den Bereich A–F beschlossen und begonnen. Die Neubearbeitung hat für die bisher erarbeiteten Alphabetabschnitte unter schwierigen äußeren Bedingungen ganz überwiegend hervorragende Ergebnisse erbracht, aber auch die Neubearbeitung wird (einschließlich der Materialsammlung) über 50 Jahre in Anspruch nehmen und doch die künftige Brauchbarkeit des Ganzen nicht sichern können. Mit den bisherigen Methoden werden wir – allein aus Gründen der erforderlichen Bearbeitungszeiten – niemals eine akzeptable Gesamtbearbeitung erhalten. Die Schaffung eines umfassenden und ausgereiften lexikographischen Standardwerks, das für das Deutsche dieselbe Funktion besitzen könnte wie die schon bestehenden Werke für das Englische, Französische, aber auch für das Dänische, das Niederländische oder das Schwedische, ist unter den in Deutschland seit langem herrschenden lexikographischen Arbeitsbedingungen nicht möglich und heute mit den Konzepten der Vergangenheit auch nicht mehr verantwortbar. Die unseren Möglichkeiten und Notwendigkeiten wirklich gerecht werdende Darstellung des deutschen Wortgebrauchs auf der Grundlage der historischen Textzeugnisse erfordert eine einschneidende Änderung der Verfahrensweisen.

Auch wenn das Eingeständnis bitter ist, wir haben uns damit abzufinden, daß die Neubearbeitung des DWB ein Torso bleibt. Wer Genaueres über Stichwörter aus dem H (Abschluß 1876), dem I (Abschluß 1877), dem K (Abschluß 1873), dem L (Abschluß 1881), dem M (Abschluß 1885), dem N (Abschluß 1885), dem O (Abschluß 1886), dem P (Abschluß 1889), dem R (Abschluß 1892) wissen will (oder über wichtige Teilbereiche auch der meisten anderen Buchstaben von G bis Z), muß sich mit dem begnügen, was unsere Urgroßväter dazu hinterlassen haben. Dies sind die Tatsachen, und dies bleibt (bis auf den sich weiter vergrößernden Abstand der Generationen) für die absehbare Zeit genau so, wie es ist.

3. Aufbau elektronischer Korpora

Die heute anwendbaren Techniken zur Einspeicherung großer Textmengen in elektronische Korpora und deren Erschließung durch geeignete Abfragesoftware, erzwingen neue Basislösungen. Die Erarbeitung lexikographischer Standardwerke in Buchform, nach denen zweifellos weiterhin Nachfrage besteht, muß durch die Verfahren der EDV ergänzt und, was weit wichtiger ist, fundiert und in bestimmten Bereichen in absehbarer Zeit auch ersetzt werden.

Je umfangreicher und je kompakter ein lexikographisches Projekt heute ist, desto dringender muß geprüft werden, ob die Darbietung des Produkts in der klassischen Buchform genügt, ob die Buchform durch eine CD-ROM zu ergänzen ist oder ob der angestrebte Zweck und das angesprochene Publikum die volle Einbeziehung einer elektronisch gespeicherten und abrufbaren Angebotsform nahe legt.

Gerade für die bisherigen Sorgenkinder der Lexikographie, die akademischen Großunternehmen, die am Tropf der öffentlichen Mittelzuweisungen dahinkümmern, stellt sich die Frage nach elektronischen Existenzformen mit besonderer Dringlichkeit: Wollen wir uns weiterhin mit der gewohnten Realisierung dieser Unternehmen in den kleinen Schritten der vom Alphabet vorgegebenen Lieferungsfolge begnügen (wonach bei glücklichem Ausgang vielleicht unsere Enkel den Abschluß der von uns begonnenen Werke erleben), oder ergänzen und ersetzen wir einen wichtigen Teil des gegenwärtigen Leistungsangebots durch die Möglichkeit der Abfrage öffentlich zugänglicher lexikographischer Datenbanken. Wenn wir uns hierfür entschließen (und dieser Entschluß will gut bedacht und mit den Interessen der Geldgeber, Verlage und mutmaßlichen Nutzer abgestimmt sein), brauchen wir, so schnell es geht, wissenschaftliche Arbeitsstellen, die große Textkorpora einspeichern, verwalten, erschließen – einschließlich der hierzu notwendigen Interpretationsleistungen – und eine Abfragesoftware, die den schnellen Zugriff auf Wörter, Wortverbindungen und Belegtexte, aber auch auf Begriffe und Strukturen aus Texten in beliebigen, also auch historischen Schreibungen ermöglicht und so sofort die Nutzung solcher Korpora von A–Z erlaubt. Die Geduldssprobe einer jahrzehntelangen Wartezeit, wie sie die großen Wörterbücher bisher regelmäßig erfordern, entfällt. Alles Eingespeicherte kann ohne Zeitverlust intern und extern abgefragt werden und dies auch schon in noch provisorischen Bearbeitungsstufen. Wir müssen uns wieder daran gewöhnen, daß der Lexikograph nicht Ewigkeitswerke verfaßt, sondern an offener Straße arbeitet, so gut er es kann, mit vollem Irrtumsrisiko.

Anzustreben ist die gleichzeitige Abrufbarkeit wichtiger Auskünfte für das gesamte Stichwortalphabet am Bildschirm in einer nach festzulegenden Kriterien gestaffelten und vom jeweiligen Bearbeitungsstand abhängigen Informationstiefe. Eine Grundmenge an Informationen (Belegsuche für Einzelwörter und Wortgruppen, Ergebnisse von chronologisch geschichteten Frequenzzählungen u.ä.) wäre schon sehr bald durch die vorhandene Software für das Gesamtalphabet automatisch darstellbar. Durch die Kombination eigener und fremder Analyseschritte würden abrufbare Auskünfte über syntaktische und semantische Grundstrukturen je nach dem fortschreitenden Entwicklungsstand der Software automatisch, halbautomatisch und durch Nacharbeit der Projektbeteiligten fortlaufend ausgebaut werden.

Die Pflege des Korpus wird zur gemeinsamen Aufgabe von Lexikographen und Informatikern. Die Erweiterung und Heranführung der Korpusbasis an die jeweilige Gegenwart ist als interne Modernisierung jederzeit neben der

laufenden lexikographischen Analyse möglich. Die Abfragesoftware kann ebenfalls ohne jede Störung der Korpora ständig verbessert werden. Auch Materialreduzierungen auf aussagefähige Teilmengen, die der knappen Information dienen, werden kein technisches Problem mehr bilden. Verfahren, die die Arbeit mit in sich gegliederten und einzeln oder in Gruppen aktivierbaren Korpuskonfigurationen erlauben, stehen in ausgereifter Form zur Verfügung. Der schnelle Zugriff auf Wörter, Wortkombinationen, Formulierungen wird auch für große Textmengen ohne weiteres möglich sein. Wenn – zumindest nach Abschluß der jetzigen Übergangsphase – das bisherige Prinzip der *alphabetgeleiteten* Korpuserschließung in einer Lieferungsfolge (nach der Methode: 500000 Stichwörter von A–Z in 100 Jahren) ersetzt wird durch das Prinzip *problemgeleiteter* Korpuserschließung bei grundsätzlicher Gleichberechtigung aller Alphabeteile, werden wir endlich Zusammengehöriges, nach Wichtigkeitsrängen geordnet, koordiniert bearbeiten können und – in dieser Rangordnung des Wichtigen – sehr viel erfolgreicher als bisher den notwendigen Kontakt zum linguistischen und lexikographischen Forschungsumfeld halten (bis hin zur verabredeten Nutzung der Forschungsergebnisse anderer in der ganzen Erstreckung des Stichwortalphabets). Gleichzeitig wird auch eine neue Offenheit für die Wünsche und Anregungen aller Interessenten deren Einfluß auf die Abfolge und die Intensität der zu leistenden Teilarbeiten ermöglichen und so das Leistungsangebot der historischen Wortforschung wieder in das Bewußtsein der Öffentlichkeit zu rücken helfen.

Mit den bisherigen Problemen der Wörterbucharbeit vertraute Lexikographen pflegen mit großer Sorge auf die Umfänge der heute erzeugbaren elektronischen Korpora zu blicken und suchen nach Möglichkeiten, die Fülle des Materials auf repräsentative, aber praktikable (in der lexikographischen Arbeit beherrschbare) Teilmengen einzuschränken. Daß Auswahlverfahren nötig sind, ist unbestritten. Fraglich ist, an welcher Stelle der Prozedur sie stattfinden sollen. Folgende Alternative stellt sich: Man kann die für ein bestimmtes lexikographisches Projekt wichtigen Belegtexte schon in der Phase der Textaufnahme herauslesen und ein Wörterbuchkorpus nur aus diesen vorgeprüften und sozusagen abgezählten Belegen bilden, die eine im Vorfeld ausreichend genau definierte Stichwortliste bedienen. Oder man baut ein offenes Korpus auf einer definierten Textbasis auf, für das Prozeduren entwickelt werden, die alle Belege für eine bestimmte Anfrage im nachhinein beherrschbar machen. Beides ist heute möglich. Ein eher statisches Korpus der erstgenannten Art wird allerdings für freie, nicht vorgeplante Anfragen wesentlich weniger aussagefähig sein als ein eher dynamisches Korpus des zweiten Typs.

Ein solches Korpus soll Anfragen beantworten können, die die deutsche Gegenwartssprache und – zumindest – ihre jüngere Vergangenheit betreffen, Fragen mit denen man sich bisher in der Regel vergeblich an wissenschaftliche Wörterbücher wendet: Wann und in welchen Zusammenhängen sind Wörter wie *Evakuierung*, *Kommunikation*, *körperbehindert*, *Umsiedler* oder

Werkstätiger zuerst aufgekomen. Seit wann waren die Naziformeln »*Arbeit macht frei*«, »*Führer befehl, wir folgen*« oder »*Heim ins Reich!*« in Gebrauch? Wie lauten die (deutschen) Frühbelege für *politische Korrektheit*, *totaler Krieg* oder *konservative Revolution und christlicher Sozialismus*? Seit wann sind Genitivfügungen wie *Haus der Jugend*, *Tag der Arbeit*, *Theater des Volkes* modisch geworden? In welcher Zitiertradition stehen Formeln wie »*einmal Polizist, immer Polizist*«, »*einmal angeschmiert, immer angeschmiert*«, »*Hiroshima ist überall*«, »*Tschernobyl ist überall*«, oder seit wann sind Titelformulierungen des Typs »*Der mit dem Wolf tanzt*« üblich geworden? Vor allem soll aber ein solches Korpus offen und dynamisch im Sinne steter Ergänzung und Aktualisierbarkeit sein. Man braucht sich nur vorzustellen, welchen Schatz wir besäßen, wenn die ungeheure Exzerptionsleistung für das DWB von etwa 1840 bis heute in ein zeitlos brauchbares Gesamtkorpus eingegangen wäre, das außerdem immer neue Abfragemodalitäten zuließe. Der Übergang vom Zettelarchiv zum elektronischen Korpus kann schon jetzt in allen Wörterbuchstellen, in denen die Druckvorbereitung über Disketten erfolgt (und das ist heute der Normalfall), über die Einspeicherung der Druckdisketten in ein eigenes Basiskorpus erfolgen. Allein durch diesen Schritt wird die Abrufbarkeit aller in sämtlichen Belegen eines Alphabetabschnitts enthaltenen Wörter für A–Z ermöglicht.

Nun mag man fragen, was die Beantwortung solcher eher klassischer Fragestellungen an Wortschatzsammlungen denn für die eingangs formulierten, eher sprachpolitischen fünf Punkte erbringen. Natürlich haben wir uns vor lexikologischer Hochstapelei zu hüten, aber sprachpolitische Probleme haben relativ viel mit Texten und den darin getroffenen Entscheidungen über Wortgebrauch und Formulierungswahl zu tun. Um Fragen dieses Typs überhaupt seriös behandeln zu können, müssen endlich auch deutsche (französische, englische usw.) Gesetzes-, Verwaltungs- und Politikertexte in der nötigen Zahl und Vielfalt elektronisch gespeichert und abrufbar erschlossen werden. Die Einspeicherung belletristischer Texte muß durch ein ausreichendes Sortiment von Gebrauchstexten ergänzt werden. Zeugnisse des innerstaatlichen Aufeinanderprallens verschiedener Sprachen, der sprachlichen Auseinandersetzung ihrer Träger und des sprachlichen Ausgleichs sind zu erfassen. Dasselbe gilt für den internationalen Sprachaustausch und für die besonderen sprachlichen Verhältnisse zwischen den Staaten mit deutscher Landessprache. Ohne einen großzügigen neuen Ansatz werden wir auch hier kaum über Sprachpolitik alten Stils hinauskommen.

Um die neuen Möglichkeiten der Korpuslexikographie im gemeinsamen Interesse verantwortlich nutzen zu können, sollten öffentliche wissenschaftliche Institutionen, an denen mit öffentlichen Mitteln erarbeitete und elektronisch gespeicherte Textkorpora bestehen, bereit sein, ihre Korpora auch für die Zwecke anderer öffentlicher wissenschaftlicher Institutionen zu öffnen. Für den Bereich der historischen Lexikographie erfordert das die Begründung eines Geistes der Kooperation zwischen allen (oder zunächst allen größeren)

korpusführenden Stellen, unabhängig davon, ob sie computerlexikographische Möglichkeiten schon nutzen oder erst künftig nutzen wollen. Aber auch andere könnten helfen, so z.B. die Deutsche Forschungsgemeinschaft, wenn sie die Förderung von Editionen grundsätzlich an die Bedingung knüpfte, die erarbeiteten Texte bei Eignung und Bedarf für den Aufbau eines elektronischen Textkorpus zur Verfügung zu stellen.

Die Begründung von Kooperationsbeziehungen zwischen unterschiedlichen Projekten ist keine Absage an bestehende Wörterbuchunternehmen, sondern die Einladung zur Bündelung der Kräfte. Zwischen den tragenden Institutionen und den Korpusverantwortlichen der Projekte sind Absprachen zu treffen über die Formen einer Zusammenarbeit, die die gegenseitige Ergänzung der Potentiale ermöglicht und Doppelarbeit, z.B. das Einscannen derselben Texte, künftig vermeiden hilft. Hierher gehören auch Verabredungen über die Korpusplanung (Grundkorpora, Spezialkorpora, Korpuskombinationen), vor allem aber über die Kompatibilität der Verarbeitungsprozeduren und der Abfragesoftware.

Eine Einigung aller interessierten Institutionen (nach Möglichkeit unter Einbeziehung wenigstens der Interessenten in Österreich und der Schweiz) sollte in gleicher Weise den Prinzipien der Zusammenarbeit im Bereich der historischen Textkorpora und großer gegenwartssprachlicher Korpora gelten. Wenn öffentliche wissenschaftliche Institutionen im gesamten Bereich des Aufbaus wissenschaftlich nutzbarer Textkorpora in abgestimmter Weise vorgehen, wäre eine wichtige Voraussetzung dafür gegeben, um endlich eine Einigung mit dem Verlagsbuchhandel über Rechtsfragen zu erreichen, die gegenwärtig den Korpusaufbau behindern. Die gering entwickelte Bereitschaft der meisten Verlage, eigene Verlagserzeugnisse zur Aufnahme in wissenschaftliche Korpora freizugeben, die über öffentliche Netze abfragbar sind, sollte durch verabredete Mißbrauchssicherungen gestärkt werden. Allgemeine juristische Lösungen (gesamteuropäische oder globale) werden kaum abgewartet werden können. Nötig ist die Einigung auf Techniken, die die berechtigten Interessen der Verlage, die ungeschützte Verbreitung ihrer Erzeugnisse zu verhindern, berücksichtigen. Nötig ist aber auch, das Interesse der Sprachwissenschaft an elektronischer Verarbeitung neuerer, noch den Copyright-Fristen unterliegender Texte zur Geltung zu bringen. Techniken, die den Erwartungen beider Seiten gerecht werden, könnten wohl schon heute verabredet und durchgesetzt werden, wenn man sich nur zur Zusammenarbeit entschließt.

Die wichtigen Fragen der Größenordnung angestrebter Lösungen, der Einigung auf technische Parameter, die Entscheidung über die einzusetzende Hardware und Software, die Art der Kooperation mit den bestehenden einschlägigen Forschungsinstitutionen, die Entscheidung für eines der verfügbaren elektronischen Lexika als Rückgrat für die Aufbereitungsprozeduren großer Textkorpora, die Entscheidungen über die Finanzierung, die Anbindung und personelle Besetzung eines neuen Korpuszentrums bleiben zunächst offen. Hier soll nur für die Einsicht geworben werden, daß neue Lösungen

dringend erforderlich, aber auch möglich sind, und brauchbare, materialgestützte Entscheidungen über einen Teil der eingangs formulierten Fragen von solchen Lösungen abhängen.

Was SARA für das British National Corpus, für Großbritannien und die ganze Anglophonie als ein Auskunftsmittel für den modernen englischen Sprachgebrauch zu leisten verspricht oder was der Internet-Server SILFIDE des Centre de Recherche en Informatique de Nancy bei der Nutzung von Sprachdaten für Frankreich und die Frankophonie an öffentlicher Brauchbarkeit erwarten läßt, ist für den deutschen Sprachbereich noch zu organisieren. So anregend Idee und Praxis des großen nationalen Deutschen Wörterbuchs im 19. Jahrhundert auf die Nachbarstaaten gewirkt haben, heute haben wir von ihnen zu lernen, wie man die Welt der eigenen Texte und den in ihnen gebrauchten Wortschatz nach Abschluß des DWB-Torsos erschließt und für Interessenten aller Art öffnet. Wir brauchen ein Korpus, ein Korpuszentrum zu seiner Aufbereitung und einen Abfragemodus, der nicht nur 2000 privilegierte Käufer eines Lieferungswerks zufriedenstellt, sondern eine neue Qualität von Öffentlichkeit der Wortschatzforschung erst ermöglicht. Die Göttinger Akademie hat mit ihrem Akademie-Thesaurus im Internet einen ersten wichtigen Schritt auf diesem Weg getan. Nötig ist aber eine Gemeinschaftsanstrengung und dringend erforderlich wäre es, für ein solches Unternehmen auch die Verlage zu gewinnen. Ein Datenbanksystem, das auf einer solchen Lösung basiert, wird keine isolierte lexikographische Bemühung bleiben, sondern reiche Ansatzpunkte zur wissenschaftlichen Nutzung für alle Disziplinen bieten, die mit Texten umzugehen haben.

Summary

Plea for a modern corpus-based German lexicography

There is an eminent research tradition within German lexicography; Grimm's dictionary, the most impressive achievement of this scholarly work, was soon to become the model of many similar enterprises. But not only is it largely outdated by now (most entries are based on work of the 19th century): there is generally an increasing gap in German lexicographical research between what is needed and possible, on the one hand, and what is actually achieved, on the other. Several reasons for this unsatisfactory situation are discussed; the most important among these is probably that the actual practice of all larger enterprises in this field is still dominated by methods of the 19th century. The new edition of Grimm's dictionary, which was started in the Fifties, will probably never be completed, if continued as at present. The only way to overcome this unsatisfactory situation and to approach the standards reached in other countries would be a comprehensive corpus-based lexical enterprise with highly flexible task-specific software tools.